



ELSEVIER

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

journal homepage: www.elsevier.com/locate/jecp

Theory of mind and executive function during middle childhood across cultures

Zhenlin Wang^{a,*}, Rory T. Devine^b, Keri K. Wong^b, Claire Hughes^b^a Department of Psychological Studies, The Hong Kong Institute of Education, Tai Po, New Territories, Hong Kong^b Centre for Family Research, Department of Psychology, University of Cambridge, Cambridge CB2 3RQ, UK

ARTICLE INFO

Article history:

Available online xxxx

Keywords:

Theory of mind
Executive function
Middle childhood
Cross-cultural comparison
Pedagogical experience
Emergence account

ABSTRACT

Previous studies with preschoolers have reported “East–West” contrasts in children’s executive function (East > West) and theory of mind (East < West). This cross-cultural study with two samples of older children from the United Kingdom and Hong Kong aimed to test competing accounts of these contrasts that focus on either global effects of culture or more specific effects of pedagogical experience. Both groups of children in Hong Kong outperformed the British children on executive function tasks. That is, with respect to executive function, general cultural influences appear to be salient. In contrast, compared with their U.K. counterparts, children attending local schools in Hong Kong (but not those attending British-based international schools in Hong Kong) performed poorly on age-appropriate tests of theory of mind. With respect to theory of mind, therefore, pedagogical experiences appear to be more salient than factors related to the broad contrast between individualist and collectivist cultures. Our findings also contribute to the debate surrounding the relationship between theory of mind and executive function; although scores on these two sets of tasks were robustly correlated within each country, the double dissociation between delayed theory of mind but superior executive function for children in local schools in Hong Kong compared with their U.K. peers suggests that variation in executive function may be necessary but is not sufficient to explain variation in theory of mind.

© 2015 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

* Corresponding author.

E-mail address: zlwang@ied.edu.hk (Z. Wang).

<http://dx.doi.org/10.1016/j.jecp.2015.09.028>

0022-0965/© 2015 The Authors. Published by Elsevier Inc.

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Introduction

Theory of mind is a developmental achievement that emerges early in life and continues to develop during adolescence and adulthood. Developments in cognitive domains such as language and executive function, as well as social factors such as cultural practice, family context, and interactional and pedagogical experience, all support the process of gaining insight into people's mental world (for a comprehensive review, see [Hughes & Devine, 2015](#)). Research in this field has been largely restricted to young children (e.g., [Wellman, Cross, & Watson, 2001](#); [Wellman & Liu, 2004](#)), although recent years have seen growing interest in infants' mental state understanding (e.g., [Onishi & Baillargeon, 2005](#); [Sodian, 2011](#)). Moreover, several studies have reported striking individual differences in adults' perspective taking ([Dodell-Feder, Lincoln, Coulson, & Hooker, 2013](#); [Ferguson & Austin, 2010](#); [Keysar, Lin, & Barr, 2003](#); [Rozzman, Cassidy, & Baron, 2003](#)).

What remains scarce, however, is research on theory of mind during middle childhood that bridges the gap between these two fields. A few studies (e.g., [Banerjee, Watling, & Caputi, 2011](#); [Devine & Hughes, 2013](#); [Dumontheil, Apperly, & Blakemore, 2010](#)) have shown age-related gains in theory of mind across middle childhood, but individual differences in theory-of-mind development during this period and their underlying mechanisms remain poorly understood. To address this challenge, the current study examined both internal and external factors that might contribute to individual difference in theory of mind beyond early childhood. With regard to internal factors, we focused on executive function, the higher level cognitive ability that underpins flexible, goal-directed activity using working memory, attention shifting, and inhibitory control ([Blair, Zelazo, & Greenberg, 2005](#); [Diamond, 2013](#)). With regard to external factors, we focused on social and pedagogical experiences at school, which increase in frequency and complexity across middle childhood ([Eccles, 1999](#)).

Theory of mind across cultures

Although most theory-of-mind research has been conducted within Anglo-Saxon countries ([Hughes & Devine, 2015](#)), the past decade has seen a marked increase in cross-cultural research (e.g., [Callaghan et al., 2005](#); [Liu, Wellman, Tardif, & Sabbagh, 2008](#)), including several cross-cultural comparisons ([Ahn & Miller, 2012](#); [Hughes et al., 2014](#); [Lecce & Hughes, 2010](#); [Lewis et al., 2009](#); [Sabbagh, Xu, Carlson, Moses, & Lee, 2006](#)). This cross-cultural perspective is helpful in unraveling the nature versus nurture riddle in theory-of-mind development: Is theory of mind an innate, culturally universal construct that is merely triggered by environmental factors, or is it cultivated in a context of social interaction, displaying culturally specific developmental routes?

Existing findings are mixed. Some cross-cultural studies (e.g., [Callaghan et al., 2005](#); [Oberle, 2009](#)) have highlighted synchrony in the onset of false belief understanding, but others have reported dramatic contrasts. For example, [Mayer and Träuble \(2013\)](#) found that under the age of 8 years Samoan children overwhelmingly failed false belief tasks, with one third of 10- to 13-year-old Samoan children also failing. These results echo earlier reports of delays in non-Western children's understanding of false belief (e.g., [Naito & Koyama, 2006](#); [Vinden, 1996](#)). Moreover, a meta-analysis of data from more than 3000 children from mainland China and Hong Kong showed that although mainland Chinese children were more or less in line with North American children in the onset of false belief understanding, children in Hong Kong lagged behind by up to 2 years ([Liu et al., 2008](#)).

If preschoolers in Hong Kong lag behind their Western peers by up to 2 years in their false belief understanding ([Liu et al., 2008](#)), do they catch up eventually? There is some evidence to suggest that Chinese adults might in fact have better perspective taking than their Western counterparts. [Wu and Keysar \(2007\)](#) found that bilingual Chinese American adults outperformed European Americans in perspective taking. However, a later reanalysis of Wu and Keysar's data using a more time-sensitive approach found that the Chinese participants made the same egocentric mistakes initially in interference as the European American participants but suppressed the interference earlier and more effectively ([Wu, Barr, Gann, & Keysar, 2013](#)). That is, rather than having a more advanced perspective taking ability, the bilingual Chinese participants appeared to capitalize on their relatively advanced executive functions ([Sabbagh et al., 2006](#)). To date, no study has gone beyond the preschool years

in comparing the social understanding of children in Hong Kong and Western children. The first aim of our study was to compare older children from the United Kingdom and Hong Kong on a battery of age-appropriate theory-of-mind tests in order to assess whether children in Hong Kong “catch up” in theory of mind during middle childhood.

Explaining cultural contrasts: The role of educational experiences

How might potential differences in theory-of-mind use between children from Hong Kong and children from the United Kingdom be explained? According to one hypothesis, our cultural norms (collectivist vs. individualist) about actions and cultural specific epistemologies shape our mental state inferences (Ames et al., 2001). For example, the emphasis on obedience and behavioral inhibitory control in parenting practice within collectivist cultures might lead to reduced conversational exposure to mental state terms (e.g., Mayer & Träuble, 2013). Related to this view, Lu, Su, and Wang (2008) found that, in both a longitudinal study and a training study, Chinese preschoolers' success on false belief tasks was associated with increased conversational references to other people rather than to talks about mental states in particular. In contrast, Chasiotis, Kiessling, Hofer, and Campos (2006) argued that conflict inhibition, but not delay inhibition, is a culture-independent, universal developmental prerequisite for the development of theory of mind.

As Hughes et al. (2014) have noted, attributing the cross-cultural difference in theory of mind to individualist versus collectivist contrasts is an oversimplification. Moreover, this hypothesis fails to account for the discrepancy between Hong Kong children and the mainland Chinese children found by Liu et al. (2008). In similar collectivist cultures, Japanese children passed the false belief task at around 6 to 8 years of age (Naito & Koyama, 2006), considerably later than the Western norm of 4 years (Wellman et al., 2001), whereas Korean children have been found to outperform their U.S. and British counterparts on false belief tasks (Ahn & Miller, 2012).

Likewise, contrasts have been reported between children from different parts of Europe: the United Kingdom and Italy (Lecce & Hughes, 2010). In the first cross-cultural study of theory of mind to establish measurement invariance (i.e., to ensure that group contrasts did not simply reflect spurious measurement effects), Hughes et al. (2014) compared means on a latent theory-of-mind factor in school-aged children from the United Kingdom, Italy, and Japan matched on age, gender, and verbal ability. Their findings replicated both the U.K.–Italy contrast (Lecce & Hughes, 2010) and previous reports of a delay in Japanese children (Naito & Koyama, 2006). However, there was no clear difference between Italian and Japanese children, challenging a simple individualist versus collectivist contrast. Instead, Hughes et al. (2014) proposed that the differences reflected contrasts in pedagogical experience because children in the United Kingdom start formal schooling at age 4 or 5 years, at least a year earlier than Italian and Japanese children.

The “pedagogical experience” hypothesis rectifies the oversimplification of individualist versus collectivist contrast by shifting the focus from a broad cultural construct to a more practical, socially organized activity (Ratner, 1999), in this case, schooling, as the primary cultural influence on psychology. This pedagogical experience hypothesis is in line with an environmental account that emphasizes the social origins of individual differences in understanding of mind. Evidence for this view comes from several distinctive lines of research, including studies of twins and deaf children born to hearing versus deaf parents, training studies, and longitudinal studies of relations between family discourse and theory of mind (for a comprehensive review, see Hughes & Devine, 2015). Formal schooling immerses children in learning activities that involve interpreting both epistemic mental states (e.g., knowledge, memories, ignorance, false belief) and motivational mental states (e.g., attention, intention). These experiences are likely to benefit children's theory-of-mind development (Frye & Wang, 2008; Tomasello, Kruger, & Ratner, 1993). Consistent with this view, theory of mind has been related to children's understanding of the concept of teaching and learning (Frye & Ziv, 2005; Wang, 2010; Ziv & Frye, 2004; Ziv, Solomon, & Frye, 2008) as well as children's own teaching activity (Davis-Unger & Carlson, 2008a, 2008b).

Our study provided a unique opportunity to test the pedagogical hypothesis. As a former British colony in the Far East that is now “Asia's World City,” Hong Kong includes both international schools, most of which adopt inquiry-based curricula and use English as a mode of instruction, and

Chinese-style local schools, which focus heavily on academic learning (Watkins & Biggs, 2001), even in early childhood classrooms. Teachers in Hong Kong local schools emphasize behavioral control and the need to follow instructions instead of inquiry (Cheng, Benson, Lau, & Fung, 2009). Children's early pedagogical experience is mainly directed to the mastery of language and literacy, which is predominantly taught in local Hong Kong schools via a drill-and-practice approach (Li & Rao, 2005) due to the fact that children live in a trilingual (Cantonese, Mandarin, and English) biliterate (Chinese and English) society. According to the pedagogical hypothesis, children attending international schools in Hong Kong should perform just as well on theory-of-mind tasks as their British counterparts because they share a similar schooling experience. In contrast, children attending local schools in Hong Kong have quite distinct school experiences that are less rich in opportunities for discussing different points of view and so might be expected to lag behind their British counterparts in their theory-of-mind performance.

The second goal of the current study was to test this pedagogical hypothesis. By pairing up U.K. pupils and Hong Kong international school pupils, as well as U.K. pupils and Hong Kong local school pupils, we singled out the effect of pedagogical experience that is usually embedded in a broader cultural context. According to the collectivist versus individualist culture divide hypothesis, both groups of children in Hong Kong should differ from children in the United Kingdom in their theory-of-mind performance. According to the pedagogical hypothesis, however, children in the United Kingdom are expected to show better theory of mind than children in Hong Kong attending local schools but not those attending international schools.

Executive function in theory of mind during middle childhood

Executive function is closely associated with the acquisition of false belief during early childhood. The relation between theory of mind and executive function can be summarized by either the "expression" account or the "emergence" account (Moses, 2001). According to the expression account, the executive demands of theory-of-mind tasks (in particular false belief tasks) mask children's understanding of mind. Specifically, children's failure on theory-of-mind tasks reflects the executive demands in those tasks rather than a lack of theory of mind. According to the emergence account, however, executive function is a necessary condition for the acquisition of theory-of-mind understanding. Children need to acknowledge different points of view and be able to hold back their own perspective in order to appreciate other perspectives. In a recent meta-analysis based on 102 studies representing close to 10,000 3- to 6-year-old children from 15 countries, Devine and Hughes (2014) reported a consistent and moderate correlation ($r = .38$) between false belief understanding and executive function that remained significant when effects of age and verbal ability were controlled. Further analysis of 10 longitudinal studies indicated that early individual differences in executive function modestly predicted later theory of mind, even when controlling for previous theory-of-mind scores and verbal ability, but not vice versa. Interestingly, the correlation between executive function and theory of mind appeared to be consistent across different cultures. The findings support an emergence account of the relation between theory of mind and executive function.

The expression account posits that children with the greatest levels of executive function performance should outperform their peers on measures of theory of mind, whereas the emergence account is more tolerant to a disassociation between theory-of-mind performance and executive function performance. Sabbagh et al. (2006) compared American and Chinese preschool children on measures of theory of mind and executive function. Despite a marked advantage in executive function relative to their American peers, Chinese children did not outperform their American counterparts on measures of theory of mind. The results seem to support the emergence account. The expression account also predicts that only the theory-of-mind tasks with high executive demands, but not those with low executive demands, would correlate with executive function; on the contrary, the emergence account predicts that executive function should correlate with theory-of-mind tasks with either high executive demands or low executive demands. Testing this possibility, Carlson, Claxton, and Moses (2015) measured preschool children with executive function tasks, theory-of-mind tasks with high executive demands, and theory-of-mind tasks with low executive demands. They found that executive function tasks, specifically conflict executive function tasks, correlated uniformly with theory-of-mind

measures that imposed either high or low executive demands when verbal ability was controlled, again supporting an emergence account.

During middle childhood and adolescence, children continue to develop both executive function (e.g., Davidson, Amso, Anderson, & Diamond, 2006; Huizinga, Dolan, & van der Molen, 2006) and theory of mind (e.g., Devine & Hughes, 2013; Dumontheil et al., 2010). Moreover, individual differences in theory of mind remain correlated with executive function during middle childhood (e.g., Bock, Gallaway, & Hund, 2014; Lagattuta, Sayfan, & Blattman, 2010; Lagattuta, Sayfan, & Harvey, 2014). To date, however, the links between theory of mind and executive function during middle childhood have not been studied in a cross-cultural context. Examining the relations between theory of mind and executive function in different cultures provides an opportunity to understand the nature of these links. Our third goal, therefore, was to compare executive function and theory-of-mind performance in school-aged children from the United Kingdom and Hong Kong using online theory-of-mind measures and examine the association between individual differences in these two domains in each sample. In doing so, we sought to examine the universality of the relation between theory of mind and executive function during middle childhood.

Moreover, our study gave us an opportunity to test the expression account versus emergence account of the executive function and theory-of-mind relation. According to the expression account, if children in Hong Kong outperformed the British children on executive function measures, they should also do better in theory-of-mind measures. According to the emergence account, children in Hong Kong would not necessarily outperform their British counterparts in theory-of-mind measures even if they did so in executive function measures. In contrast to false belief tasks, in which children need to inhibit their own more salient knowledge about a situation in order to infer a protagonist's false belief, the online measures of theory of mind simply require children to explain the protagonist's behaviors and so impose a lower executive demand. According to the expression account, these measures would not correlate well with executive function measures. The emergence account predicts that executive function measures should correlate with theory-of-mind measures with either high executive demands or low executive demands (Moses, 2001).

In summary, our study had three key aims: (a) to investigate similarities and differences in theory of mind in children from the United Kingdom and Hong Kong, (b) to assess the role of pedagogical experiences in shaping individual differences in theory of mind, and (c) to examine the universality of the relations between theory of mind and executive function during middle childhood. To address these three aims, we drew on two data sets that included aggregate measures of theory of mind during middle childhood that were administered to samples of children in Hong Kong and the United Kingdom.

Method

Participants

Sample 1 consisted of 118 children (48% male) recruited from seven English-speaking international schools in Hong Kong and six state primary and secondary schools in the United Kingdom as part of a follow-up study of children's social development in the United Kingdom and Hong Kong led by the third author (Wong, Freeman, & Hughes, 2014). Follow-up children were matched in age, $t(118) = 0.57$, $p = .57$, and self-reported family affluence, $t(117) = 0.49$, $p = .63$. Inclusion criteria for both sites were (a) no known developmental delays or disabilities and (b) native speaker of English or spoke English as a second language. The U.K. sample was predominantly White British (77.5%, $n = 39$; 53% male) with a mean age of 12.42 years ($SD = 1.89$, range = 9.00–16.07). The Hong Kong sample was more ethnically diverse ($n = 78$; 45% male) than the U.K. sample (77.5% White British) in that children reported being Chinese (43.6%), White British (17.9%), and mixed race (17.9%) with a mean age of 12.38 years ($SD = 2.00$, range = 9.10–15.62). However, a t -test confirmed that the two samples did not differ in age, $t(116) = 0.10$, $p = .92$. In both samples, English was the primary language spoken at home (United Kingdom = 97.5%; Hong Kong = 61.5%), with Cantonese (16.7%) and Mandarin (9%) being the two other most spoken languages at home. In terms of family background, the children from the United Kingdom

had significantly more siblings ($M = 1.37$, $SD = 0.88$), $t(101) = 2.20$, $p < .05$, $d = 0.45$, than the children from Hong Kong ($M = 1.01$, $SD = 0.72$). Parental education levels were higher in the Hong Kong sample than in the U.K. sample, $\chi^2(1) = 9.89$, $p < .01$, $\phi = .31$. Specifically, 68 of 78 mothers in the Hong Kong sample had a university degree or higher, whereas 33 of 39 mothers in the U.K. sample had upper secondary school education or a university degree. In terms of socioeconomic status, 80.3% to 85.4% of U.K. and Hong Kong participants fell in the “affluent” band, respectively, as measured by the Family Affluence Scale (Boyce, Torsheim, Currie, & Zambon, 2006).

Sample 2 consisted of 137 children taking part in the fifth wave of an ongoing longitudinal study of social and cognitive development in the United Kingdom (Hughes, 2011). We recruited a comparable sample of 125 children recruited from state primary schools in Hong Kong. Inclusion criteria for both sites were (a) no known developmental delays or disabilities and (b) native speaker of English/Cantonese (as appropriate). From this sample of 262 children, we created two groups of children individually matched on age (in months) and gender from the Hong Kong sample to children in the U.K. sample. In each site, there were 108 children (57% male). The U.K. sample was predominantly White British with a mean age of 10.81 years ($SD = 0.39$, range = 10.05–11.55). The Hong Kong sample was predominantly ethnic Chinese with a mean age of 10.81 years ($SD = 0.35$, range = 10.05–11.54); a t -test confirmed that the two samples did not differ in age, $t(214) = -0.05$, $p = .96$. In terms of family background and socioeconomic status, the children from the United Kingdom had significantly more siblings ($M = 2.02$, $SD = 1.45$), $t(214) = 6.34$, $p < .001$, $d = 0.87$, than the children from Hong Kong ($M = 1.01$, $SD = 0.80$). Parental education levels were also higher in the U.K. sample than in the Hong Kong sample, $\chi^2(1) = 16.75$, $p < .001$, $\phi = .28$. Specifically, whereas 64 mothers in the U.K. sample had upper secondary school education (e.g., A levels) or a university degree, only 33 of the mothers in the Hong Kong sample had upper secondary school education or a university degree. In terms of self-rated affluence, children in the United Kingdom ($M = 6.09$, $SD = 1.80$) gave themselves higher ratings on the Family Affluence Scale than children in Hong Kong ($M = 4.46$, $SD = 1.86$), $t(212) = 6.53$, $p < .001$, $d = 0.90$. These between-sample differences in family background and socioeconomic status were controlled statistically in each of our models.

Measures

Table 1 summarizes the measures administered to Samples 1 and 2.

Theory of mind

Each theory-of-mind task involved watching a short film clip or listening to a short audio file (~30 s in length) and then either explaining a character's behavior or answering short questions about a character's thoughts and feelings. The children from both samples completed the Triangles Task (Castelli, Happé, Frith, & Frith, 2000). In this task, the children watched three short silent cartoons depicting two triangles moving about against a white background. Designed to elicit mental state attributions, these cartoons featured instances of coaxing, teasing, and surprising. Using the coding scheme developed by

Table 1
Measures administered to Samples 1 and 2.

Domain	Measure	Sample 1	Sample 2
Theory of mind	Triangles Task	✓	✓
	Silent Film Task	✓	✓
	Strange Stories	×	✓
Executive function	Bead Memory	✓	✓
	Digit Span Backward	✓	×
	Trail Making	✓	×
	Arrows Task	×	✓
	Smiling Faces	×	✓
Verbal ability	Word Reasoning (WISC-IV)	✓	×
	BPVS	×	✓

Castelli et al. (2000), the children's responses were scored for ascription of intentionality (i.e., the proclivity to ascribe mental states when explaining the actions of the triangles) and accuracy (i.e., how closely descriptions matched the intended story). Intentionality scores ranged from 0 (non-deliberate action) to 5 (deliberate actions with the goal of affecting others' mental states). Accuracy was scored as 0 (incorrect descriptions), 1 (imprecise descriptions), or 2 (correct descriptions of the story presented in the clip). We created total scores by summing together intentionality and accuracy scores for each item (possible range = 0–21).

The children in both samples also completed the Silent Film Task (Devine & Hughes, 2013). In this task, the children were required to explain the behavior of characters in five clips (~30 s in length) from a classic silent comedy played once in a fixed order. The first clip was accompanied by two questions, and the remaining four clips were followed by one question. Following the coding scheme developed by Devine and Hughes (2013), participants received 2 points for responses that correctly described the events shown in the clip with reference to characters' mental states, 1 point for answers that correctly described the events but did not use mental explanations, and 0 points for irrelevant or factually incorrect responses. Scores for individual items were summed together to create a total score (possible range = 0–12).

In addition, the children in Sample 2 completed the Strange Stories Task. For this task, we used five vignettes (taken from Happé, 1994) that depicted social situations involving double bluffs, deception, and misunderstanding. After each vignette, the children were asked to explain a character's behavior. The vignettes were translated into Cantonese by a panel of three English/Cantonese bilingual developmental psychologists that included the lead author, adopting a collaborative and iterative translation approach to ensure conceptual equivalence (Douglas & Craig, 2007). In each site, children listened to recordings (~30 s in length) of an adult reading each of the stories. The text of each story remained on-screen after the audio recording stopped. Following the coding scheme developed by White, Hill, Happé, and Frith (2009), fully correct responses received 2 points, partially correct responses received 1 point, and incorrect responses received 0 points; across the five vignettes, therefore, children could score between 0 and 10 points.

The validity of these tasks as measures of theory of mind is supported by three sources of evidence. First, these tasks have shown concurrent associations in a large sample of children ages 7 to 13 years even when individual differences in verbal ability and narrative comprehension were taken into account (Devine & Hughes, 2015). Second, performance on these tasks at age 10 years has also been shown to be linked to earlier performance on a battery of false belief understanding tasks at age 6 years (Devine, White, Ensor, & Hughes, submitted for publication). Third, performance on these tasks has been shown to be related to individual differences in children's self-reported social competence (Devine & Hughes, 2013).

For each task, the children's responses were digitally recorded and later transcribed verbatim for coding. To ensure comparability of ratings, 25% of the transcripts in each site in both studies were double-coded. Individual items exhibited moderate to strong levels of reliability. In Sample 1, the two-way mixed single measure model type with absolute accuracy showed excellent average measure reliability for the Silent Film Task total score (intraclass correlation (ICC) [3, 1] = .96, 95% confidence interval (CI) [.89, .98], $p < .001$). The kappa values for all items were within acceptable ranges ($\kappa = .73$ – 1.00 , $SE = .00$ – $.12$). Similarly, the total intentionality (TTI) and appropriateness (TTA) ratings on the Triangles Task showed excellent average measures reliability (TTI: ICC[3, 1] = .96, 95% CI [.91, .98]; TTA: ICC[3, 1] = .90, 95% CI [.78, .96]; $ps < .001$), with item-level ICCs ranging from .76 to .92. In Sample 2, the second author trained the Hong Kong team in the rating procedures using transcripts from the U.K. sample as well as translated transcripts from the Hong Kong sample. For the Strange Stories Task, the mean κ was .92 and total summed scores showed excellent levels of interrater reliability (ICC = .97, 95% CI [.94, .99], $p < .001$). For the Triangles Task, the mean κ was .77 and the total summed scores showed excellent levels of interrater reliability (ICC = .97, 95% CI [.94, .99], $p < .001$). Finally, the individual items of the Silent Film Task showed acceptable levels of interrater reliability, the mean κ was .82, and summed total scores showed excellent interrater agreement (ICC = .94, 95% CI [.87, .97], $p < .001$).

Executive function

The children in both samples completed the Bead Memory Task from the Stanford–Binet Intelligence Scale (Thorndike, Hagen, & Sattler, 1986). On each trial, the participants were shown (for 5 s) a picture of an arrangement of beads on a stick (across trials, the beads varied in number, shape, color, and position). The participants were then given a box of beads and were asked to reproduce each bead arrangement exactly. The task was discontinued after three failures across four trials. Raw scores were calculated by subtracting the number of failed trials from the highest item attempted. Although initially designed to measure short-term visual memory, the Bead Memory Task also measures multiple executive functions. Specifically, participants must (a) refrain from touching the beads while the image is displayed, (b) hold the image in mind while they select the beads, and (c) plan ahead in order to place the beads on the stick in the correct sequence. Performance on this task is correlated with performance on measures of inhibitory control and set shifting during childhood (e.g., Hongwanishkul, Happaney, Lee, & Zelazo, 2005; Hughes & Ensor, 2011).

The children in Sample 1 also completed the Digit Span Backward Test and the Trail Making Test. The Digit Span Backward Test (Wechsler Intelligence Scale for Children–Fourth Edition [WISC-IV]; Wechsler et al., 2004) is a widely used assessment of working memory. The examiner read out a list of numbers at approximately one digit per second, and the participants were asked to repeat this list of digits in reverse order. Each list was read out once and increased by one digit until the participants failed two consecutive lists of the same digit span. The correct responses were summed to create a total score (ranging from 0 to 16) and standardized scores. Scores for the current study ranged from 4 to 16. The Trail Making Test A and B (Corrigan & Hinkeldey, 1987) is a two-part test measuring switching and visual attention. The participants were timed on how quickly they could connect circles without lifting their pen off the page. This was first done sequentially on a page of 25 randomly scattered “numbered” circles on Form A (1-2- . . . 25) and then using an alternating order of “numbered–lettered” circles on Form B (1-A-2-B, etc.). Each form started with practice items, and mistakes were corrected and factored into total task completion time. The time difference between Form B and Form A provided an index of executive function, where a longer time (in seconds) reflected poorer performance. In this study, the time difference between forms ranged from 0 to 180 s. This was standardized, and a higher value reflected poorer performance.

The children in Sample 2 completed the Arrows Task (Davidson et al., 2006), a measure of inhibitory control. In this task, participants viewed one of four images of a purple arrow on a white screen and needed to respond by hitting a key on the left (1) or right (0). The arrow pointed either directly downward or diagonally on either the left- or right-hand side of the screen. During control trials, the arrow appeared on either the left or right and pointed directly downward, and participants needed to press the button on the same side as where the arrow appeared. During test trials, the arrow appeared on either the left or right but pointed diagonally. On these trials, participants needed to press the button on the opposite side of where the arrow appeared. The participants received detailed instructions and practice items prior to completing the task. The arrows were displayed for up to 750 ms and were preceded by a 500-ms interval in which a crosshair was displayed against a white background. The children completed 12 control trials and 12 test trials in a random order. To measure inhibitory control, we calculated efficiency scores based on the total number of correct test trials divided by the total time taken on test trials.

The children in Sample 2 also completed the Smiling Faces Task (Huizinga et al., 2006), a measure of cognitive flexibility. During this task, the children needed to respond with a button press to one of four cartoon faces (i.e., a happy boy, a happy girl, a sad boy, and a sad girl) displayed in one of four quadrants on a white screen. In single task trials, the children needed to identify whether the face was a boy or a girl if it appeared in the top two quadrants (by pressing 1 or 2) or to identify whether the face was happy or sad if it appeared in the bottom two quadrants (by pressing 3 or 4). During alternating trials, the children needed to integrate both rules. The trials were administered in four blocks in a fixed order: one set of 16 single task trials (boy or girl), another set of 16 single task trials (happy or sad), and two sets of 16 alternating trials. The trials within each block were presented in a random order. The participants were provided with detailed instructions and were asked to repeat back the rules of the task before each block. Each trial lasted up to 3500 ms, during which time the participants needed to respond with a button press. The trials were preceded by a 500-ms interval in which a black

crosshair was displayed in the center of a white screen. We calculated efficiency scores based on the total number of correct alternating trials divided by the total time taken to complete the alternating trials.

Verbal ability

Unfortunately, there was no measure of verbal ability that had been standardized for use in both Hong Kong and the United Kingdom. The children in Sample 1 completed the Word Reasoning Test from the WISC-IV (Wechsler et al., 2004). In this task, the participants needed to identify the concept being described in a series of 24 clues of increasing difficulty. Each item was scored as either correct (1) or incorrect (0) before five consecutive “blanks” produced a total verbal ability score out of 24. We standardized raw summed scores (with a possible range of 0–24) within each country using T-scores so that each child’s verbal ability was measured against peers in his or her own country.

For the children in Sample 2, we adopted the British Picture Vocabulary Scale (BPVS; Dunn & Dunn, 2007) because this is a widely used task that has very simple instructions. On each trial, the examiner read aloud a word and the children needed to point to one of four pictures that provided the best match for that word. We used the stimulus booklets and vocabulary lists from the U.K. version of this task. Adopting a back-translation approach (Brislin, 1970), each word was translated into Cantonese and then back-translated into English by a panel of three English/Cantonese bilingual developmental psychologists that included the lead author. The panel discussed and modified the Cantonese version to ensure that the two versions were equivalent in meaning. The Cantonese version was checked against “Lexical items with English explanations for fundamental Chinese learning in Hong Kong schools” provided by the Hong Kong Education Bureau (2009) to ensure progressive difficulty of the vocabulary. We calculated raw scores by subtracting the number of errors made by each participant from the item number corresponding to the last number in the participant’s ceiling set (i.e., the set containing eight or more errors). We then standardized scores within each country using T-scores so that each child’s verbal ability was measured against peers in his or her own country.

Procedures

The children in Sample 1 completed a 60-min individual testing session at school with the third author on two theory-of-mind tasks (Silent Film and Triangles), three executive function tasks (Beads, Digit Span Backward, and Trail Making), and a language test, among other tasks. All tasks were administered in English in both the United Kingdom and Hong Kong. At each site, the children in Sample 2 completed individual test sessions in a quiet environment at school. During these sessions (which lasted ~90 min, including breaks), an experienced graduate researcher administered a battery of tasks presented in a counterbalanced order in either English (in the United Kingdom) or Cantonese (in Hong Kong). The U.K. team provided the Hong Kong team with detailed training and observed pilot sessions in Hong Kong to ensure that the data collection procedures were identical.

Results

Analytic approach

In analyzing the data from both samples, we used confirmatory factor analysis (CFA) to test measurement models designed to reduce the number of variables in our data and provide error-free parameter estimates. Next, we used multiple indicators, multiple causes (MIMIC) models (Brown, 2006) to examine cross-cultural differences in both the theory-of-mind and executive function latent factors. Given the presence of non-normally distributed variables in each data set, we estimated each model using robust maximum likelihood estimation (as opposed to maximum likelihood estimation) in Mplus 7 (Muthén & Muthén, 2012). The robust maximum likelihood estimator provides more accurate parameter estimates than standard maximum likelihood estimation (Brown, 2006; Kline, 2011). For each model, we evaluated fit using Brown’s (2006) four recommended criteria: a nonsignificant chi-square (χ^2) test, comparative fit index (CFI) $\geq .90$, Tucker–Lewis index (TLI) $\geq .90$, and root mean

square error of approximation (RMSEA) $\leq .08$. The effect sizes for the various model parameters were interpreted in accordance with recommendations from Kline (2011); small standardized effect sizes ranged from .10 to .30, moderate effect sizes ranged from .30 to .50, and large effect sizes were greater than .50.

Descriptive statistics and data reduction

Table 2 shows descriptive statistics (and intercorrelations) for each of the key variables for Sample 1. There was no evidence of ceiling or floor effects on either the Triangles Task or Silent Film Task. There were no differences between boys and girls in age, self-reported family affluence, or performance on any measures of verbal ability, theory of mind, or executive function ($-1.36 \leq t \leq 1.33$, all $ps > .10$). There were modest correlations between the two measures of theory of mind. The three measures of executive function were moderately intercorrelated. With the exception of the Trail Making Task and the Digit Span Backward Task, each of the variables conformed to normality (i.e., $Z_{\text{skew}} < \pm 3.29$). Table 3 shows descriptive statistics (and intercorrelations) for each of the key variables for Sample 2. Each of the three measures of theory of mind captured a wide range of individual differences, as evidenced by the symmetrical distribution of scores on each task. There was no evidence of ceiling or floor effects on either the Triangles Task or Silent Film Task; no participants scored 0, and only 1% of the participants obtained the maximum score on either task. The Strange Stories Task showed some evidence of negative skew, with 12.5% of the participants achieving the maximum score. With the exception of the Arrows Task efficiency score, Strange Stories Task total score, and BPVS score, all of the key variables conformed to normality (i.e., $Z_{\text{skew}} < \pm 3.29$).

Data reduction

In Sample 1, we tested a two latent factor model using CFA. First, we loaded both theory-of-mind indicators onto a single theory-of-mind latent factor. Next, we loaded each of the three executive function indicators onto a separate (but correlated) executive function latent factor. This measurement model provided a good fit to the data, $\chi^2(4) = 1.60$, $p = .81$, RMSEA = .00, CFI = 1.00, TLI = 1.00, Akaike

Table 2
Descriptive statistics and correlations for key variables in Sample 1.

	Age	Verbal	Triangles Task	Silent Film Task	Bead Memory	Digit Span Backward	Trail Making (B–A)
Age (years)	–	.28*	.41**	.06	.47**	.38**	.42**
WISC-IV verbal (raw score)	.32*	–	.08	.07	.29*	.29**	.35**
Triangles Task	–.01	.29	–	.39**	.21	.12	.17
Silent Film Task	.32*	.19	.13	–	.07	–.02	.12
Bead Memory	.28	.35*	.38*	.26	–	.45**	.33**
Digit Span Backward	–.13	.13	.44**	.23	.39*	–	.47**
Trail Making (reversed)	.50**	.06	.11	.28	.44**	.12	–
Mean U.K.	12.42	13.01	12.79	8.10	27.58	7.83	–43.74
H.K.	12.39	13.37	12.90	7.91	28.54	8.58	–38.19
SD U.K.	1.89	3.97	4.23	2.53	3.88	2.06	28.27
H.K.	2.00	3.46	5.00	2.17	3.84	2.34	27.66
Range U.K.	9.00–16.07	4–20	1–20	0–12	21–36	5–14	–122.8 to –32.26
H.K.	9.10–15.62	5–20	0–21	0–12	21–37	4–16	–148.10 to –3.29

Note. U.K., United Kingdom; H.K., Hong Kong; SD, standard deviation.

* $p < .05$.

** $p < .01$.

Table 3

Descriptive statistics and correlations for key variables in Sample 2.

	Age	BPVS	Strange Stories	Triangles Task	Silent Film Task	Bead Memory	Arrows	Smiling Faces
Age (years)	–	.22*	.15	.20*	–.05	.18	.15	.19*
BPVS (raw score)	.17	–	.58**	.29**	.38**	.52**	.32**	.35**
Strange Stories	.06	.23*	–	.35**	.34**	.43**	.28**	.31**
Triangles Task	.18	.29**	.30**	–	.22*	.18	.17	.11
Silent Film Task	.07	.13	.22*	.42**	–	.21**	.10	.09
Bead Memory	.24*	.36**	.32**	.25**	.19*	–	.31**	.34**
Arrows (trials/s)	–.01	.24*	.17	.24*	.05	.13	–	.32**
Smiling Faces (trials/s)	.27**	.24*	.04	.14	.15	.25**	.20*	–
Mean U.K.	10.81	134.50	6.52	12.53	7.63	26.02	0.67	0.44
H.K.	10.81	148.69	7.39	12.69	6.77	27.48	0.61	0.51
SD U.K.	0.34	12.93	2.45	3.77	2.13	3.56	0.48	0.12
H.K.	0.33	7.37	2.01	4.12	2.15	3.49	0.53	0.16
Range U.K.	10.05–11.54	98–163	1–10	3–19	1–12	18–36	0–1.82	0.10–0.93
H.K.	10.05–11.55	101–160	2–10	3–21	1–12	19–35	0–2.40	0.15–0.90

Note. U.K., United Kingdom; H.K., Hong Kong; SD, standard deviation. Correlations for the U.K. sample are presented above the diagonal. Correlations for the H.K. sample are presented below the diagonal.

* $p < .05$.

** $p < .01$.

information criterion (AIC) = 3463.10. The standardized factor loadings for the theory-of-mind latent factor were .75 and .38 ($ps < .01$). The standardized factor loadings for the executive function latent factor ranged from .63 to .70 (all $ps < .001$). The theory-of-mind latent factor was moderately correlated with the executive function latent factor ($\phi = .43$, $p = .01$).

In Sample 2, we specified a two latent factor model in which each of the three theory-of-mind task indicators loaded onto a theory-of-mind latent factor and each of the three executive function task indicators loaded onto an executive function latent factor. The two latent factor model fit the data well, $\chi^2(8) = 13.64$, $p = .09$, RMSEA = .05, CFI = .95, TLI = .91, AIC = 7146.43. Both latent factors accounted for significant variance in task performance. The three theory-of-mind indicators loaded significantly onto the theory-of-mind latent factor (mean loading = .50, range = .30–.69, all $ps < .001$). The three executive function indicators loaded significantly onto the executive function latent factor (mean loading = .50, range = .38–.67, all $ps < .001$). The two latent factors were strongly correlated ($\phi = .77$, $p < .001$). Note that we chose to use a two latent factor solution for two reasons. First, from a conceptual perspective, theory of mind and executive function are related but distinct constructs. Meta-analytic evidence shows that individual differences in executive function and theory-of-mind task performance are only moderately correlated (Devine & Hughes, 2014). Moreover, findings suggest that the “real-life” correlates of executive function and theory of mind are distinct; for example, the relations between executive function and problem behaviors are significantly greater than those between theory of mind and problem behaviors in preschool children (Hughes & Ensor, 2008). Second, from an empirical standpoint, an alternative single latent factor model in both samples did not provide a good fit to the data (Sample 1: $\chi^2(5) = 8.96$, $p = .11$, RMSEA = .08, CFI = .94, TLI = .88, AIC = 3467.98; Sample 2: $\chi^2(8) = 18.24$, $p = .03$, RMSEA = .07, CFI = .92, TLI = .87, AIC = 7149.03).

Modeling theory-of-mind task performance

Using data from Sample 1, we examined whether there were cultural differences in theory-of-mind task performance using a structural equation in which the two latent factors of theory of mind and executive function were regressed onto a binary “nation” variable. To reduce the potential influence

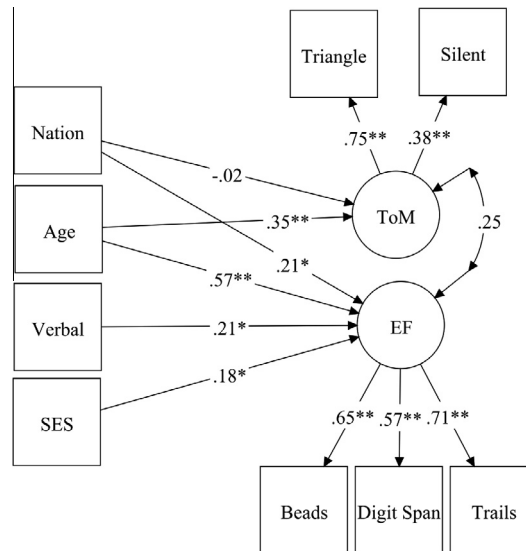


Fig. 1. Robust maximum likelihood estimates for correlates of individual differences in theory of mind (ToM) and executive function (EF) in Sample 1. * $p < .05$; ** $p < .01$.

of confounding variables, we also regressed each latent factor onto three continuous variables: age, verbal ability, and self-reported family affluence (the latter of which is a socioeconomic status [SES] indicator). This model provided an excellent fit to the data, $\chi^2(16) = 19.23$, $p = .26$, RMSEA = .04, CFI = .97, TLI = .95, AIC = 3423.38. The parameter estimates for this model are presented in Fig. 1. When accounting for individual differences in executive function, age, verbal ability, and family affluence, there were no significant differences between the British children and the Hong Kong international school children in their performance on the theory-of-mind latent factor. In contrast, there was a weak but significant difference in performance between children in Hong Kong and children in the United Kingdom on the executive function latent factor, with children from Hong Kong outperforming their British counterparts by .21 standardized units. This difference was independent of individual differences in age, verbal ability, theory of mind, and family affluence.

Next, using data from Sample 2, we specified a similar structural equation model to examine cultural differences in theory-of-mind task performance. To minimize potential confounds, we also regressed each latent factor onto age, gender, number of siblings, parental education, and verbal ability scores. Our initial model did not provide a good fit to the data, $\chi^2(32) = 68.41$, $p < .001$, RMSEA = .07, CFI = .86, TLI = .78, AIC = 6888.13. Inspection of the modification indices revealed evidence of differential item functioning in that nation exerted a direct effect on performance on both the Strange Stories Task and Triangles Task (see below for more details). The inclusion of these regression paths improved the overall model fit, $\chi^2(30) = 37.13$, $p = .17$, RMSEA = .03, CFI = .97, TLI = .95, AIC = 6860.85 (see Fig. 2). Together, these models accounted for 65% of the variance in theory-of-mind latent factor scores and 64% of the variance in executive function latent factor scores.

Echoing findings from preschool samples (e.g., Sabbagh et al., 2006), scores on the theory-of-mind latent factor in Sample 2 were significantly lower for children from Hong Kong than for children from the United Kingdom. Even when potential differences in executive function, parental education, number of siblings, and verbal ability were controlled, children from Hong Kong were, on average, .60 standardized units lower than children from the United Kingdom, indicating a medium to large difference in performance (Brown, 2006; Cohen, 1988). The direct effect of nation on the Strange Stories Task and Triangles Task indicators (independent of individual differences in the theory-of-mind latent factor) showed that these indicators were not culturally invariant. That is, performance on these two tasks

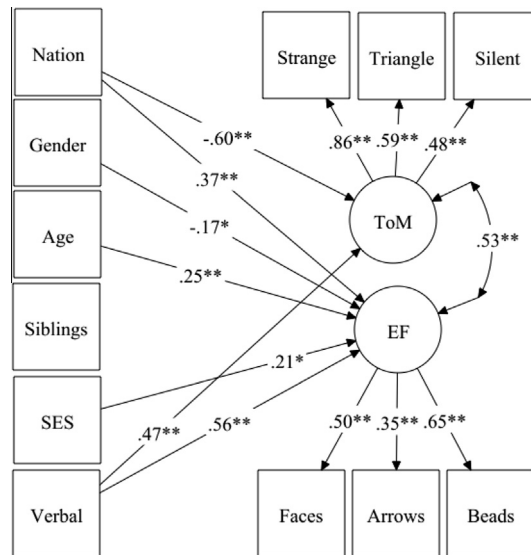


Fig. 2. Robust maximum likelihood estimates for correlates of individual differences in theory of mind (ToM) and executive function (EF) in Sample 2. $^{*}p < .05$; $^{**}p < .01$.

differed between both sites for reasons other than theory of mind. In sum, there were national differences favoring children in Hong Kong in aspects of these tasks that did not related to mental state reasoning.

In direct contrast to the results for the theory-of-mind latent factor, children from Hong Kong outperformed children from the United Kingdom on the executive function latent factor. When effects of theory of mind, parental education, number of siblings, and verbal ability were controlled, children from Hong Kong obtained scores that were, on average, .37 standardized units higher than those for children from the United Kingdom. This is equivalent to a small difference in performance (Brown, 2006). There were no direct effects of nation on the indicators of executive function, suggesting that these items were culturally invariant.

The structural equation model revealed moderate to strong correlations between theory of mind and executive function latent factor scores as well as between each of these latent factors and verbal ability. Parental education, SES, age, and gender exerted small but significant effects on executive function (but not on theory of mind). The number of siblings a child had was unrelated to either theory of mind or executive function.

Discussion

In this article, we have presented data from two cross-cultural comparisons that jointly produced three main findings. First, when effects of general child and family characteristics were taken into account, theory-of-mind latent factor scores were equivalent for children in the United Kingdom and children attending international schools in Hong Kong, but children attending local schools in Hong Kong scored, on average, .60 standardized units lower than children from the United Kingdom, indicating a medium to large effect. Second, the contrast between children in Hong Kong attending local schools and those attending international schools was specific to theory of mind; both groups outperformed their British counterparts on the executive function latent factor with small effect sizes, .23 standardized units in Sample 1 and .37 standardized units in Sample 2. Third, in both samples, the latent factors for theory of mind and executive function were correlated with either a moderate or large effect size.

Our motivation in conducting this research was to discover whether children in Hong Kong “catch up” in theory-of-mind performance during middle childhood. The results revealed no evidence for such a catch-up, at least for Hong Kong local school pupils. The findings extend earlier reports of young Hong Kong children’s delay on theory of mind (Liu et al., 2008) and demonstrate a persistent lag in Hong Kong children’s social understanding. Furthermore, this lag could not be explained by factors related to the global contrast between individualist and collectivist cultures because, unlike their peers from Hong Kong local schools, children from Hong Kong international schools performed just as well as their British counterparts on theory of mind. Instead, the results favor the pedagogical experience hypothesis, namely that children who were exposed to inquiry-based pedagogy (children in the United Kingdom and children attending Hong Kong international schools) showed better theory-of-mind performance than children who were exposed to the drill-and-practice pedagogy (Hong Kong local school children).

What makes this conclusion more convincing is the fact that children attending international schools in Hong Kong were better on executive function than the U.K. children, just as children attending local schools in Hong Kong outperformed their British counterparts. In other words, Hong Kong international school pupils were culturally distinct from the U.K. children. Hong Kong international school students were more likely to be bilingual, a factor that is believed to facilitate theory-of-mind performance through enhanced attention control and inhibition rather than conceptual mental state understanding per se (Bialystok & Senman, 2004; Kovács, 2009). By including executive function in our models, we also controlled the potential confounding effect of bilingualism. Cantonese was the main mode of instruction in Hong Kong local schools, in contrast to English in Hong Kong international schools. Among other differences, the two languages differ in syntactical complement, which is hypothesized to affect theory-of-mind development (de Villiers & de Villiers, 2000). However, findings from a cross-lingual study (Cheung et al., 2004) suggest that syntax of complement per se does not contribute to theory-of-mind development; rather, it is the general language comprehension that matters.

Our study contributes to the ongoing debate about the cultural universality versus specificity of theory-of-mind development by suggesting that it is not prudent to attribute cross-cultural differences in theory-of-mind development to a global individualist versus collectivist cultural distinction. Children’s direct social environments, their micro systems in Bronfenbrenner’s (1977) terms, might play a more important role in shaping the diverse paths of their theory-of-mind development. The pedagogical experience hypothesis offers a more specific mechanism accounting for individual differences in theory of mind and highlights the importance of the quality of education in social understanding development. However, further longitudinal and intervention studies are necessary to establish a causal relation between pedagogical experiences and theory-of-mind development.

The strong correlation between theory of mind with low executive demands and executive function in both samples extends Devine and Hughes’s (2014) meta-analytic finding regarding the association between these two constructs during early childhood into middle childhood. The universality of the relation across an extended period in development and across cultures provides a basis for understanding the nature of the relation between these constructs. Hong Kong children’s advantage in executive function and concurrent disadvantage in theory of mind during middle childhood resonate with an earlier report of a similar inconsistency during early childhood (Sabbagh et al., 2006). This finding challenges a simple “expression” account and supports an “emergence” account of the relation between theory of mind and executive function. Viewed alongside the results from a meta-analysis of longitudinal data (Devine & Hughes, 2014), our cross-cultural data indicate that during middle childhood, just as during the preschool years, executive function facilitates (but is distinct from) theory of mind.

Limitations

The two samples were originally recruited for different studies. Sample 1 was slightly older than Sample 2, and the children in the two samples completed slightly different tasks. Therefore, it is challenging to compare Hong Kong international school students with the local school students directly.

Future studies are needed to explore the individual differences in mental state understanding among different populations within Hong Kong.

Another potential limitation in our study is that there were socioeconomic status contrasts between the different samples. Compared with the Hong Kong pupils at local schools in Sample 2, the Hong Kong international school participants in Sample 1 were from more affluent families. Within Sample 2, the Hong Kong participants were from less affluent households with less educated parents when compared with their British counterparts. We statistically controlled for these variables in all of the models to account for these differences. It is worth noting that, with respect to parental education, Sample 2 was representative of the respective populations. In comparison with the 30% to 50% of employed U.K. adults who have higher education qualifications (Higher Education Statistics Agency, 2013), higher education in Hong Kong is available for only approximately 20% of the population (Hong Kong Education Bureau, n.d.). This contrast supports the ecological validity of any conclusions drawn from these samples.

Conclusion

This study is the first East versus West comparison of theory of mind and executive function during middle childhood. By expanding the age range to the much neglected period of middle childhood, our study adds valuable evidence on theory-of-mind use during this stage of development. Building on the well-documented delay in Hong Kong preschoolers' theory-of-mind acquisition, our study demonstrated, for the first time, the similarities and differences in mental state understanding beyond early childhood in children in Hong Kong when compared with children in the United Kingdom. Specifically, the contrast between children in Hong Kong attending local schools and those attending international schools carries important educational implications.

Our study found that children in Hong Kong attending local (but not international) schools showed a delay in theory-of-mind development compared with their British counterparts. In contrast, both groups of children from Hong Kong outperformed the British children on executive function measures. These results highlight the potential cost of drilling and rote learning (the dominant model in local Hong Kong schools) for children's understanding of others. Our study also extends earlier reports on the association between theory of mind and executive function of preschoolers to middle childhood and demonstrates that, despite a clear advantage on executive function tasks, children in Hong Kong do not outperform their British counterparts on tests of theory of mind. This dissociation suggests an interesting contrast in the salience of social influences on executive function and theory of mind that deserves further examination in future studies.

Acknowledgments

This study was funded by the U.K. Economic and Social Research Council and Hong Kong RGC Joint Research Scheme Project ES/J021180/1.

References

- Ahn, S., & Miller, S. A. (2012). ToM and self-concept: A comparison of Korean and American children. *Journal of Cross-Cultural Psychology*, 43, 671–686.
- Ames, D. R., Knowles, E. D., Rosati, A. D., Morris, M. W., Kalish, C. W., & Gopnik, A. (2001). The social folk theorist: Insights from social and cultural psychology on the contents and contexts of folk theorizing. In B. Malle, L. Moses, & D. Baldwin (Eds.), *Intentions and intentionality: Foundations of social cognition* (pp. 307–329). Cambridge, MA: MIT Press.
- Banerjee, R., Watling, D., & Caputi, M. (2011). Peer relations and the understanding of faux pas: Longitudinal evidence for bidirectional associations. *Child Development*, 82, 1887–1905.
- Bialystok, E., & Senman, L. (2004). Executive processes in appearance–reality tasks: The role of inhibition of attention and symbolic representation. *Child Development*, 75, 562–579.
- Blair, C., Zelazo, P. D., & Greenberg, M. T. (2005). The measurement of executive function in early childhood. *Developmental Neuropsychology*, 28, 561–571.
- Bock, A. M., Galloway, K. C., & Hund, A. M. (2014). Specifying links between executive functioning and theory of mind during middle childhood: Cognitive flexibility predicts social understanding. *Journal of Cognition and Development*, 16(3). <http://dx.doi.org/10.1080/15248372.2014.888350>.

- Boyce, W., Torsheim, T., Currie, C., & Zambon, A. (2006). The Family Affluence Scale as a measure of national wealth: Validation of an adolescent self-report measure. *Social Indicators Research*, 78, 473–487.
- Brislin, R. W. (1970). Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology*, 1, 185–216.
- Bronfenbrenner, U. (1977). Toward an experimental ecology of human development. *American Psychologist*, 32, 513–531.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford.
- Callaghan, T., Rochat, P., Lillard, A., Claux, M., Odden, H., Itakura, S., ... Singh, S. (2005). Synchrony in the onset of mental-state reasoning: Evidence from five cultures. *Psychological Science*, 16, 378–384.
- Carlson, S. M., Claxton, L. J., & Moses, L. J. (2015). The relation between executive function and theory of mind is more than skin deep. *Journal of Cognition and Development*, 16, 186–197.
- Castelli, F., Happé, F., Frith, U., & Frith, C. (2000). Movement and mind: A functional imaging study of perception and interpretation of complex intentional movement patterns. *NeuroImage*, 12, 314–325.
- Chasiotis, A., Kiessling, F., Hofer, J., & Campos, D. (2006). Theory of mind and inhibitory control in three cultures: Conflict inhibition predicts false belief understanding in Germany, Costa Rica, and Cameroon. *International Journal of Behavioral Development*, 30, 249–260.
- Cheng, P. W. D., Benson, P., Lau, G., & Fung, K. H. C. (2009). *Play-based pedagogy in early childhood classrooms in Hong Kong and its impact on quality education*. Hong Kong: Hong Kong Institute of Education (second report).
- Cheung, H., Hsuan-Chih, C., Creed, N., Ng, L., Ping Wang, S., & Mo, L. (2004). Relative roles of general and complementation language in theory-of-mind development: Evidence from Cantonese and English. *Child Development*, 75, 1155–1170.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.
- Corrigan, J. D., & Hinkley, N. S. (1987). Relationships between Parts A and B of the Trail Making Test. *Journal of Clinical Psychology*, 43, 402–409.
- Davidson, M. C., Amso, D., Anderson, L. C., & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia*, 44, 2037–2078.
- Davis-Unger, A. C., & Carlson, S. M. (2008a). Children's teaching skills: The role of theory of mind and executive function. *Mind, Brain, and Education*, 2, 128–135.
- Davis-Unger, A. C., & Carlson, S. M. (2008b). Development of teaching skills and relations to theory of mind in preschoolers. *Journal of Cognition and Development*, 9, 26–45.
- de Villiers, J. G., & de Villiers, P. A. (2000). Linguistic determinism and the understanding of false beliefs. In P. Mitchell & K. Riggs (Eds.), *Children's reasoning and the mind* (pp. 191–228). Hove, UK: Psychology Press.
- Devine, R. T., White, N., Ensor, R., & Hughes, C. (submitted for publication). *Theory of mind in middle childhood: Longitudinal associations with executive function and social competence*.
- Devine, R. T., & Hughes, C. (2013). Silent Films and Strange Stories: Theory of mind, gender, and social experiences in middle childhood. *Child Development*, 84, 989–1003.
- Devine, R. T., & Hughes, C. (2014). Relations between false belief understanding and executive function in early childhood: A meta-analysis. *Child Development*, 85, 1777–1794.
- Devine, R. T., & Hughes, C. (2015). Measuring theory of mind in middle childhood: Reliability and validity of the Silent Films and Strange Stories tasks. *Journal of Experimental Child Psychology*. <http://dx.doi.org/10.1016/j.jecp.2015.07.011>.
- Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, 64, 135–168.
- Dodell-Feder, D., Lincoln, S. H., Coulson, J. P., & Hooker, C. I. (2013). Using fiction to assess mental state understanding: A new task for assessing theory of mind in adults. *PLoS One*, 8(11), e81279.
- Douglas, S. P., & Craig, C. S. (2007). Collaborative and iterative translation: An alternative approach to back translation. *Journal of International Marketing*, 15, 30–43.
- Dumontheil, I., Apperly, I. A., & Blakemore, S. J. (2010). Online usage of theory of mind continues to develop in late adolescence. *Developmental Science*, 13, 331–338.
- Dunn, L. M., & Dunn, D. M. (2007). *PPVT-4: Peabody Picture Vocabulary Test* (4th ed.). Circle Pines, MN: American Guidance Service.
- Eccles, J. S. (1999). The development of children ages 6 to 14. *Future of Children*, 9(2), 30–44.
- Ferguson, F. J., & Austin, E. J. (2010). Associations of trait and ability emotional intelligence with performance on theory of mind tasks in an adult sample. *Personality and Individual Differences*, 49, 414–418.
- Frye, D., & Wang, Z. (2008). Theory of mind, understanding teaching, and early childhood education. In C. Fiorello & K. Thurman (Eds.), *Cognitive development in K–3 classroom learning: Research applications* (pp. 85–104). Mahwah, NJ: Lawrence Erlbaum.
- Frye, D., & Ziv, M. (2005). Teaching and learning as intentional activities. In B. D. Homer & C. Tamis-LeMonda (Eds.), *The development of social cognition and communication* (pp. 231–258). Mahwah, NJ: Lawrence Erlbaum.
- Happé, F. (1994). An advanced test of theory of mind: Understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *Journal of Autism and Developmental Disorders*, 24, 129–154.
- Higher Education Statistics Agency. (2013). *Patterns and trends in UK higher education*. Retrieved from: <<http://www.universitiesuk.ac.uk/highereducation/Documents/2013/PatternsAndTrendsInUKHigherEducation2013.pdf>>.
- Hong Kong Education Bureau. (2009). *Lexical items with English explanations for fundamental Chinese learning in Hong Kong schools*. Retrieved from: <<http://www.edbchinese.hk/lexlist/>>.
- Hong Kong Education Bureau. (n.d.). *Distribution of educational attainment of population aged 15 and over*. Retrieved from: <<http://www.edb.gov.hk/en/about-edb/publications-stat/figures/educational-attainment.html>>.
- Hongwanishkul, D., Happaney, K. R., Lee, W. S., & Zelazo, P. D. (2005). Assessment of hot and cool executive function in young children: Age-related changes and individual differences. *Developmental Neuropsychology*, 28, 617–644.
- Hughes, C. (2011). Changes and challenges in 20 years of research into the development of executive functions. *Infant and Child Development*, 20, 251–271.
- Hughes, C., Devine, R. T., Ensor, R., Koyasu, M., Mizokawa, A., & Lecce, S. (2014). Lost in translation? Comparing British, Japanese, and Italian children's theory-of-mind performance. *Child Development Research*, 2014. <http://dx.doi.org/10.1155/2014/893492>.
- Hughes, C., & Devine, R. T. (2015). A social perspective on theory of mind. In R. M. Lerner (Ed.), *Socioemotional processes* (7th ed., In & M. E. Lamb (Eds.), *Handbook of child psychology* (7th ed.) (Vol. 3, pp. 564–609). Hoboken, NJ: John Wiley.

- Hughes, C., & Ensor, R. (2008). Does executive function matter for preschoolers' problem behaviors? *Journal of Abnormal Child Psychology*, 36(1), 1–14.
- Hughes, C., & Ensor, R. (2011). Individual differences in growth in executive function across the transition to school predict externalizing and internalizing behaviors and self-perceived academic success at 6 years of age. *Journal of Experimental Child Psychology*, 108, 663–676.
- Huizinga, M., Dolan, C. V., & van der Molen, M. W. (2006). Age-related change in executive function: Developmental trends and a latent variable analysis. *Neuropsychologia*, 44, 2017–2036.
- Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition*, 89, 25–41.
- Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York: Guilford.
- Kovács, Á. M. (2009). Early bilingualism enhances mechanisms of false-belief reasoning. *Developmental Science*, 12, 48–54.
- Lagattuta, K. H., Sayfan, L., & Blattman, A. J. (2010). Forgetting common ground: Six- to seven-year-olds have an overinterpretive theory of mind. *Developmental Psychology*, 46, 1417–1432.
- Lagattuta, K. H., Sayfan, L., & Harvey, C. (2014). Beliefs about thought probability: Evidence for persistent errors in mindreading and links to executive control. *Child Development*, 85, 659–674.
- Lecce, S., & Hughes, C. (2010). The Italian job? Comparing theory of mind performance in British and Italian children. *British Journal of Developmental Psychology*, 28, 747–766.
- Lewis, C., Koyasu, M., Oh, S., Ogawa, A., Short, B., & Huang, Z. (2009). Culture, executive function, and social understanding. *New Directions for Child and Adolescent Development*, 123, 69–85.
- Li, H., & Rao, N. (2005). Curricular and instructional influences on early literacy attainment: Evidence from Beijing, Hong Kong, and Singapore. *International Journal of Early Years Education*, 13, 235–253.
- Liu, D., Wellman, H. M., Tardif, T., & Sabbagh, M. A. (2008). Theory of mind and development in Chinese children: A meta-analysis of false-belief understanding across cultures and languages. *Developmental Psychology*, 44, 523–531.
- Lu, H., Su, Y., & Wang, Q. (2008). Talking about others facilitates theory of mind in Chinese preschoolers. *Developmental Psychology*, 44, 1726–1736.
- Mayer, A., & Träuble, B. E. (2013). Synchrony in the onset of mental state understanding across cultures? A study among children in Samoa. *International Journal of Behavioral Development*, 37, 21–28.
- Moses, L. J. (2001). Executive accounts of theory-of-mind development. *Child Development*, 72, 688–690.
- Muthén, L. K., & Muthén, B. O. (2012). *Mplus user's guide* (7th ed.). Los Angeles: Muthén & Muthén.
- Naito, M., & Koyama, K. (2006). The development of false-belief understanding in Japanese children: Delay and difference? *International Journal of Behavioral Development*, 30, 290–304.
- Oberle, E. (2009). The development of theory of mind reasoning in Micronesian children. *Journal of Cognition and Culture*, 9, 39–56.
- Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, 308, 255–258.
- Ratner, C. (1999). Three approaches to cultural psychology: A critique. *Cultural Dynamics*, 11, 7–31.
- Rozzman, E. B., Cassidy, K. W., & Baron, J. (2003). "I know, you know": Epistemic egocentrism in children and adults. *Review of General Psychology*, 7, 38–65.
- Sabbagh, M. A., Xu, F., Carlson, S. M., Moses, L. J., & Lee, K. (2006). The development of executive functioning and theory of mind: A comparison of Chinese and U.S. preschoolers. *Psychological Science*, 17, 74–81.
- Sodian, B. (2011). Theory of mind in infancy. *Child Development Perspectives*, 5, 39–43.
- Thorndike, R. L., Hagen, E. P., & Sattler, J. M. (1986). *Stanford-Binet Intelligence Scale*. Chicago: Riverside.
- Tomasello, M., Kruger, A. C., & Ratner, H. H. (1993). Cultural learning. *Behavioral and Brain Sciences*, 16, 495–511.
- Vinden, P. G. (1996). Junin Quechua children's understanding of mind. *Child Development*, 67, 1707–1716.
- Wang, Z. (2010). *Mindful learning: Children's developing theory of mind and their understanding of the concept of learning* (unpublished doctoral dissertation). University of Pennsylvania.
- Watkins, D. A., & Biggs, J. B. (2001). The paradox of the Chinese learner and beyond. In D. A. Watkins & J. B. Biggs (Eds.), *Teaching the Chinese learner: Psychological and pedagogical perspectives* (pp. 3–23). Hong Kong: University of Hong Kong, Comparative Education Research Centre.
- Wechsler, D., Kaplan, E., Fein, D., Kramer, J., Morris, R., Delis, D., et al (2004). *Wechsler Intelligence Scale for Children – fourth edition*. San Antonio, TX: Psychological Corporation.
- Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory of mind development: The truth about false-belief. *Child Development*, 72, 655–684.
- Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development*, 75, 523–541.
- White, S., Hill, E., Happé, F., & Frith, U. (2009). Revisiting the Strange Stories: Revealing mentalizing impairments in autism. *Child Development*, 80, 1097–1117.
- Wong, K. K., Freeman, D., & Hughes, C. (2014). Suspicious young minds: Paranoia and mistrust in 8- to 14-year-olds in the UK and Hong Kong. *British Journal of Psychiatry*, 205, 221–229.
- Wu, S., Barr, D. J., Gann, T. M., & Keysar, B. (2013). How culture influences perspective taking: Differences in correction, not integration. *Frontiers in Human Neuroscience*, 7. <http://dx.doi.org/10.3389/fnhum.2013.00822>.
- Wu, S., & Keysar, B. (2007). The effect of culture on perspective taking. *Psychological Science*, 18, 600–606.
- Ziv, M., & Frye, D. (2004). Children's understanding of teaching: The role of knowledge and belief. *Cognitive Development*, 19, 457–477.
- Ziv, M., Solomon, A., & Frye, D. (2008). Young children's recognition of the intentionality of teaching. *Child Development*, 79, 1237–1256.